

TOELICHTING OP DE NORMERING 2021

We kunnen dit jaar de centrale examens niet normeren zoals anders. We hanteren een aanpak die daar zo dicht mogelijk bij in de buurt komt. Daarmee zorgen we ervoor dat de normering ook dit jaar recht doet aan de prestaties van leerlingen, waardoor leerlingen terecht een voldoende of onvoldoende blijven behalen. En we zorgen ervoor dat de eisen die gesteld worden aan leerlingen zo veel mogelijk gelijk blijven aan de eisen van voorgaande jaren.

INFORMATIE VOORAF

Diverse bronnen van informatie zijn beschikbaar voor de normering:

- Allereerst natuurlijk de afnamegegevens van de deelnemende kandidaten;¹
- Aan de docenten worden ook enkele vragen gesteld:
 - a) een vraag over de representativiteit van hun deelnemende kandidaten in vergelijking met al hun kandidaten;
 - b) een oordeel over de moeilijkheid van het examen;
- Daarnaast zijn er nog historische N-termen beschikbaar.

De wijze waarop deze bronnen van informatie worden ingezet bij het bepalen van de N-termen wordt hieronder beschreven. We maken daarbij onderscheid tussen drie tijdvakken:

- tijdvak 1
- tijdvak 2
- tijdvak 3r - op de scholen

De hieronder beschreven methode is in principe² van toepassing op alle papieren en digitale centrale examens³ en de cpe's⁴. De leerlingen die een examen maken via het Staatsexamen krijgen hetzelfde examen als leerlingen die een examen maken op hun school. Ze krijgen ook dezelfde N-term. Daarmee is deze nieuwe manier van normeren automatisch ook van toepassing voor de Staatsexamenleerlingen.⁵

TIJDVAK 1

STAP 1: N-TERM OP BASIS VAN RELATIEVE NORMERING

Bij relatief normeren wordt de N-term zo ingesteld dat dit min of meer dezelfde gemiddelde cijfers oplevert als in andere jaren. Hiervoor is wel een representatieve steekproef nodig. Dat wil zeggen dat de kandidaten in de steekproef qua vaardigheid vergelijkbaar moeten zijn met alle kandidaten die in 2021 examens doen in dit vak.

Aangezien er een extra mogelijkheid voor een eerste kans (namelijk in tijdvak 2) komt, houden we er rekening mee dat er selectieve deelname van hele scholen, en binnen scholen, kan optreden. Om de representativiteit te controleren, wordt dit jaar aan de docenten in Wolf gevraagd of de kandidaten die deelnemen, representatief zijn voor al hun kandidaten. Als uit een antwoord blijkt dat een niet-representatief deel van hun eigen kandidaten heeft deelgenomen, wordt deze school uit de steekproef gelaten. Zo blijven er scholen over waarvan we de prestaties kunnen vergelijken met hun prestaties in voorgaande jaren. Hierbij houden we rekening met de historische prestaties van de kandidaten op de scholen in de steekproef.

Dit leidt tot een eerste technisch advies voor de N-term (N-stap 1) op basis van een relatieve normering.

¹ De scores van Staatsexamenleerlingen worden nooit meegenomen bij de vaststelling van de N-term en dat zal dit jaar dus ook niet gebeuren.

² Er zijn vakken waarvoor we niet over alle hieronder beschreven informatie beschikken. Dat kan bijvoorbeeld zijn omdat een vak een nieuw programma kent of omdat een vak op zeer weinig scholen wordt afgenomen. In dat geval wordt de hierna beschreven systematiek zoveel als mogelijk intact gelaten, gezien het ontbreken van de betreffende gegevens.

³ De manier waarop de N-termen van de verschillende varianten worden bepaald, verschilt niet van die in andere jaren.

⁴ Dit geldt niet voor de cspe's voor de beroepsgerichte profielvakken die dit jaar met een schoolexamen worden afgesloten.

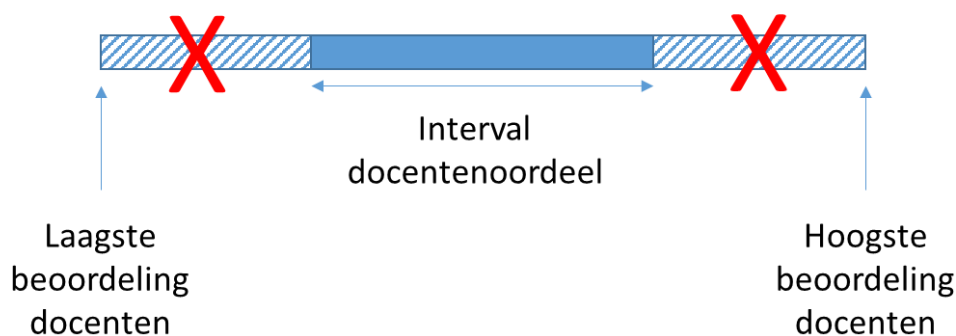
⁵ Uitzondering hierop vormt het derde tijdvak aangezien er voor staatsexamenleerlingen een apart derde tijdvak via de staatsexamencommissie is.

STAP 2: TECHNISCHE N-TERM OP BASIS VAN DOCENTENoordeel EN HISTORISCHE N-TERM

Op de N-term die volgt uit stap 1 (N-stap1) worden twee elementen toegepast die ervoor moeten zorgen dat enerzijds de prestatie-eis niet onder een bodem-niveau zakt, en anderzijds een leerling minimaal het cijfer krijgt dat hij of zij onder normale omstandigheden ook zou krijgen. Dus de normering mag niet te soepel worden, maar ook niet strenger dan onder normale omstandigheden.

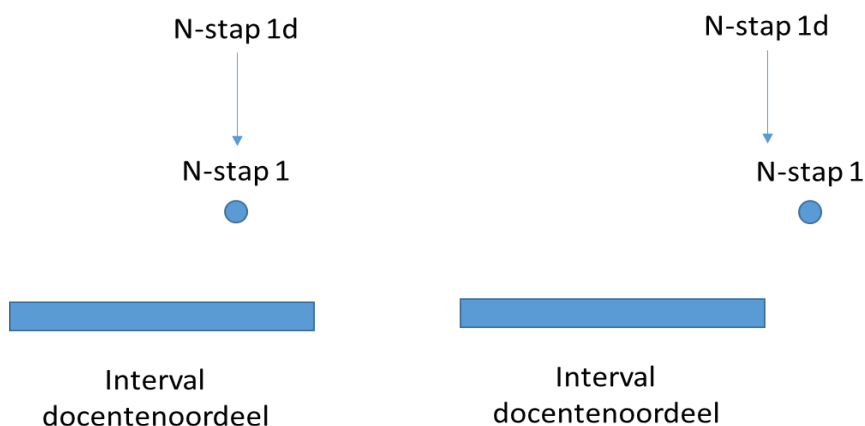
Docentenoordelen

Het eerste element zijn de docentenoordelen over de moeilijkheidsgraad van het examen. De docenten geven aan bij welke score een 5,5 behaald moet worden gezien de moeilijkheid van het examen in vergelijking met andere jaren. Deze score wordt per docent vertaald in een N-term. Dit levert een set van N-termen op. Van deze set worden de soepelste en de strengste weggestreept. De set die overblijft levert een interval van N-termen op.



Afbeelding 1: Voorbeeld interval docentenoordeel over de moeilijkheidsgraad van het examen

Vervolgens wordt nagegaan of N-stap1 binnen het interval valt. Als dat niet het geval is, wordt N-stap1 zodanig aangepast dat deze er alsnog binnen valt. De N-term die hieruit volgt noemen we N-stap1d.

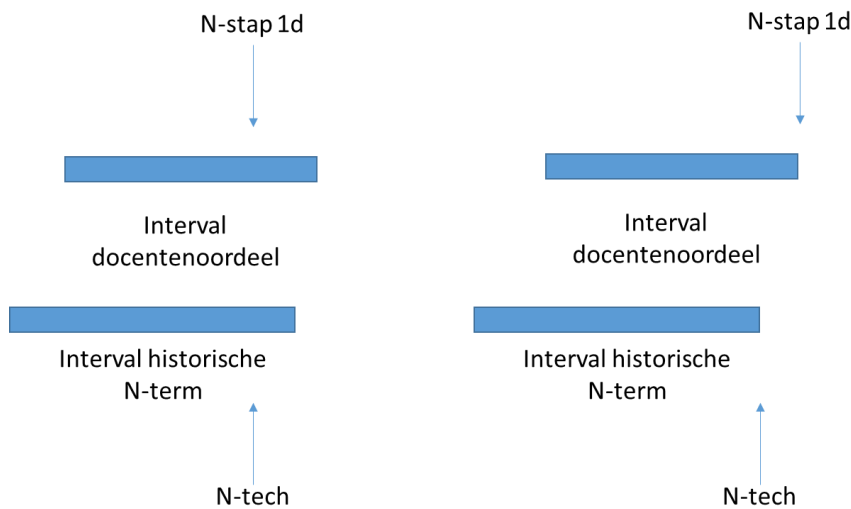


Afbeelding 2: Twee voorbeelden waarin gecontroleerd wordt of N-stap1 in het interval van docentenoordelen valt. In het linker voorbeeld is dat het geval en is aanpassing van de N-term niet nodig. In het rechter voorbeeld is aanpassing van de N-term wel nodig.

Historische N-termen

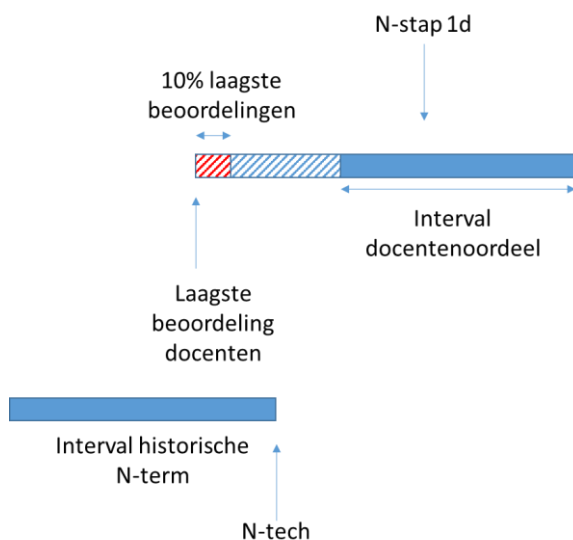
Als tweede element worden de historische N-termen betrokken bij de normering. Een interval van historische N-termen beschrijft de gangbare moeilijkheid van examens van dit vak. Het examen in 2021 wijkt naar verwachting niet dramatisch af qua moeilijkheid van voorgaande versies; de best passende N-term gezien de moeilijkheid van het examen zal dus in het interval liggen. Voor het bepalen van dit interval wordt gekeken naar een reeks recente jaren, waarin hetzelfde programma

gold en het examen niet ingrijpend veranderd is qua opbouw of moeilijkheid.⁶ Ook dit levert een interval van N-termen op. Als N-stap1d buiten het historische interval ligt, wordt de N-term bijgesteld tot de grens van het historische interval. Dit leidt tot het technisch advies voor de N-term (N-tech).



Afbeelding 3: Twee voorbeelden waarin gecontroleerd wordt of N-stap1d in het interval van de historische N-termen valt. In het linker voorbeeld is dat het geval en is aanpassing van de N-term niet nodig en in het rechter voorbeeld is aanpassing van de N-term wel nodig.

Meerdere keren is gevraagd wat we doen als het docentenoordeel helemaal niet overlapt met het interval van de historische N-term. Aangezien zowel het docentenoordeel als de historische N-term een indicator zijn van de moeilijkheidsgraad van het examen, is de kans dat ze helemaal niet overlappen bijzonder klein. Als dat wel het geval is, schuiven we in principe op naar de grens van het interval van historische N-termen. Daarbij kijken we nog wel of nog minimaal 10% van de docenten een dergelijke lagere N-term acceptabel zou vinden (zie afbeelding 4). Als er een dergelijk extreem verschil is tussen het docentenoordeel en de historische N-term dan gaan we wel na of er nog andere bronnen zijn die erop wijzen dat het examen extreem moeilijk was. We kunnen bijvoorbeeld kijken naar het oordeel van onze vakdeskundigen over de moeilijkheidsgraad van het examen. De kans dat een examen zo moeilijk is dat de N-term buiten het interval van historische N-termen valt is al klein en dat het er zo extreem buiten valt is nog kleiner. We verwachten dus niet dat deze situatie zich zal voordoen.



Afbeelding 4: voorbeeld waarbij het interval van docentenoordelen niet overlapt met het interval van historische N-termen.

⁶ In enkele gevallen kan het voorkomen dat er geen historisch interval is omdat het vak of het examen(programma) net nieuw is. In dat geval is N-stap1d meteen het technisch advies voor de N-term.

STAP 3: DEFINITIEVE N-TERM

Als laatste worden de meldingen van docenten via de Examenlijn, van leerlingen via het LAKS en van vakverenigingen beoordeeld door onze vakdeskundigen. Zij kunnen adviseren om de technische N-term nog op te hogen om te compenseren voor fouten of onvolkomenheden. Dit leidt tot de definitieve N-term.

TIJDVAK 2

In de juni-afname kunnen er zowel kandidaten zijn die voor het eerst deelnemen aan het examen (eerstekansers) als kandidaten die ook reeds aan tijdvak 1 hebben deelgenomen (herkansers). Omdat de twee groepen verschillend zijn, kun je ze niet op één hoop gooien bij het bepalen van de N-term. Om die reden kijken we eerst naar beide groepen afzonderlijk om de N-term van het examen te bepalen.

Op de scores van de 'eerstekansers' doen we dat op dezelfde manier die ook in tijdvak 1 wordt toegepast. Op de scores van de herkansers doen we dat volgens de methode die normaal gesproken in tijdvak 2 wordt toegepast.

Dat geeft ons op twee manieren informatie over de N-term van het examen in tijdvak 2. Door beide uitkomsten te combineren komen we tot de definitieve N-term. De mate waarin beide uitkomsten daarop van invloed zijn, hangt af van de grootte van beide groepen. Uiteindelijk resulteert dit in één N-term in tijdvak 2. Deze is voor alle leerlingen gelijk, ongeacht of ze 'eerstekanser' of herkanser zijn. Net zoals in tijdvak 1 wordt in de laatste stap bekeken of compensatie voor fouten of onvolkomenheden nodig is. Daarna wordt de definitieve N-term vastgesteld.

GEEN VOORLOPIGE N-TERM

In voorgaande jaren was het gebruikelijk dat er een voorlopige N-term gepubliceerd werd voor het tweede tijdvak. Deze kon dan alleen naar boven worden bijgesteld. Het publiceren van een voorlopige N-term zou voor eerstekansers in tijdvak 2 een andere uitgangssituatie opleveren dan voor leerlingen die in tijdvak 1 examen doen. Daarnaast kan een vooraf vastgestelde minimale N-term leiden tot lagere eisen in tijdvak 2 dan in tijdvak 1. Om die reden wordt in 2021 geen voorlopige N-term bekend gemaakt. De N-term van tijdvak 2 kan lager worden vastgesteld dan de N-term van tijdvak 1 als alle signalen erop wijzen dat het examen in tijdvak 2 makkelijker was dan het examen in tijdvak 1.

TIJDVAK 3R – OP DE SCHOLEN

In tijdvak 3r hebben we naar verwachting te maken met leerlingen die een herkansing maken nadat ze hun examen gemaakt hebben in tijdvak 1 of in tijdvak 2. Dat kan een verschil opleveren in de mate waarin leerlingen zich nog kunnen verbeteren. Aangezien de mate waarin leerlingen zich verbeteren een rol speelt bij het bepalen van de N-term, kun je ook de scores van deze groepen niet op één hoop gooien. De scores van beide groepen worden apart bekeken bij het bepalen van de moeilijkheidsgraad van het examen. Ook nu geldt dat het om één en hetzelfde examen gaat en dus om één moeilijkheidsgraad. Uiteindelijk levert dit dus ook één N-term op die recht doet aan de prestaties van alle leerlingen in dit tijdvak. Tot slot wordt in de laatste stap bekeken of compensatie voor fouten of onvolkomenheden nodig is. Daarna wordt de definitieve N-term vastgesteld.

GEEN VOORLOPIGE N-TERM

Ook bij de juli-afname kan er geen voorlopige N-term bekend worden gemaakt. Er zou dan namelijk gekozen moeten worden voor die van de mei-afname of die van de juni-afname. Daarnaast zou de uitgangssituatie voor herkansende leerlingen in tijdvak 2 ongelijk zijn aan die voor herkansende leerlingen in tijdvak 3r.